

DOCUMENT DE TRAVAIL

DEVISING AND USING
EVALUATION STANDARDS
THE FRENCH PARADOX

JEAN-CLAUDE BARBIER

N° 23

avril 2003



**CENTRE
D'ETUDES
DE L'EMPLOI**

«LE DESCARTES I»
29, PROMENADE MICHEL SIMON
93166 NOISY-LE-GRAND CEDEX
TÉL. 01 45 92 68 00 FAX 01 49 31 02 44
MÉL. cee@cee.enpc.fr
<http://www.cee-recherche.fr>

Devising and Using Evaluation Standards The French Paradox

JEAN-CLAUDE BARBIER

jean-claude.barbier@mail.enpc.fr

CNRS/Centre d'études de l'emploi

DOCUMENT DE TRAVAIL

N° 23

avril 2003

ISSN 1629-7997
ISBN 2-11-088618-8

Devising and Using Evaluation Standards

The French Paradox

Jean-Claude Barbier

Abstract

The developments of evaluation studies and research in France were given a new impetus in the nineties. Yet there is no such thing as a proper “evaluation profession” in this country. Actually, there exists rather a heterogeneous constellation of advocacy coalitions, with very diverging approaches to evaluation. The dominant and rather subdued conception is that the distinctive characteristics of a professional activity labelled “evaluation” are not so important indeed. Hence the difficult task of adopting even minimum standards of quality and propriety.

However, the text shows that, in the nineties, the *Conseil scientifique de l'évaluation (CSE)* promoted a very specific and interesting approach, with a particular stress put upon the conditions of designing and steering the evaluations. It is interesting to analyse how the former *CSE* has actually implemented criteria to meta-evaluate a significant body of evaluation reports, which were conducted under its scientific supervision at that time.

Key words: evaluation, public policies.

Conception et utilisation des standards d'évaluation

Le paradoxe français

Résumé

Le présent texte propose une analyse des développements de l'évaluation en France depuis les années quatre-vingt-dix ainsi que l'explication de la fragilité de cette « profession » (qui n'en est pas une), en termes d'advocacy coalition.

Les conceptions de l'évaluation, en France, sont aussi diverses que les différents milieux qui les portent : pour une part d'entre eux, il s'agit d'aboutir, sinon à une hégémonie, tout au moins à l'érection d'une conception dominante.

Le texte s'attache à analyser l'une de ces conceptions : celle qui fut construite systématiquement par le Conseil scientifique de l'évaluation. Le cœur de cette approche a reposé sur un dispositif relativement original. À l'écart des stratégies politiques et administratives, une autorité scientifique s'efforce, au nom de l'intérêt général, d'établir les conditions de possibilité d'une construction d'un jugement (sur des politiques et des programmes publics) le plus justifiable, aussi bien dans ses attendus, ses hypothèses que dans sa logique argumentative.

Mots-clefs : évaluation, politiques publiques.

Avertissement

Le présent document de travail constitue la version préliminaire d'un chapitre écrit pour un livre¹ collectif dirigé par des collègues spécialistes de l'évaluation (R. Schwartz [Israël], J. Mayne [Canada] et J. Toulemonde [France]).

Les impératifs de l'édition en langue anglaise - que je désapprouve fortement - s'imposent de plus en plus et, dans le cas de ce texte, ont conduit à éliminer l'essentiel des citations en français, de même qu'une bonne part des références bibliographiques françaises.

C'est la première raison pour laquelle, en parallèle avec l'édition future, ce texte me semble pouvoir s'insérer dans la liste des Documents de travail du Centre d'études de l'emploi. Il y a, dans l'approche française des questions d'évaluation, des spécificités qui, pour être bien comprises, réclament d'être formulées en français, même si on trouve des équivalents approximatifs en anglais, pour l'intercompréhension. Cette remarque vaut plus généralement sur le plan de la recherche internationale. La domination de la langue anglaise est préjudiciable à la qualité de la recherche, et non pas simplement pour des raisons de préservation (certes justifiée) de ce qu'il est convenu de nommer « la diversité culturelle », mais pour des raisons théoriques de fond.

Au reste, le présent texte comporte deux types d'analyse, peu répandus en France. Le premier propose une analyse des développements de l'évaluation en France depuis les années quatre-vingt-dix, et l'explication de la fragilité de cette « profession » (qui n'en est pas une), en termes d'advocacy coalition, reprenant la notion de Sabatier (1998). Le second touche au cœur des conceptions diverses de l'évaluation que, précisément, en France, plusieurs milieux différents entendent contrôler, au moins pour une part d'entre eux, afin d'aboutir, sinon à une hégémonie, du moins à l'érection d'une conception dominante. Le texte analyse l'une d'entre elles, celle qui fut construite systématiquement par feu le Conseil scientifique de l'évaluation (CSE). Le cœur de cette conception repose sur un dispositif relativement original vis-à-vis des conceptions comparables à l'étranger : à l'écart (momentané) des stratégies politiques et administratives (selon la conception de Leca, ex-président du CSE), une autorité scientifique s'efforce, au nom de l'intérêt général, d'établir les conditions de possibilité d'une construction d'un jugement (sur des politiques et des programmes publics) le plus justifiable, aussi bien dans ses attendus, ses hypothèses que dans sa logique argumentative.

¹ *Assessing Evaluative Information* est le titre provisoire de l'ouvrage, qui doit être publié en 2003 chez Transaction Publishers (Rutgers, N.J.).

INTRODUCTION

In comparative European terms, the French history of evaluating is relatively short. In 1990, it was generally thought that a breakthrough had been achieved and that evaluation had been so to say “institutionalised”².

Evaluation practice subsequently grew considerably, i.e. practice calling itself “evaluation”, in a context of continuous controversy between different conceptions and of an increasing stimulus provided both by the European Union and the process of devolution to the regional authorities (*conseils régionaux*). Nevertheless no homogenous “evaluation community” has emerged in France and no set of professional guidelines, standards or norms for assessing evaluation has ever been formally adopted by any significant group of professionals.

On the other hand, the period 1990-96 has effectively seen the dissemination and usage of norms by some institutions, among which the *Conseil scientifique de l'évaluation (CSE)* was instrumental and prominent; this is somehow paradoxical.

After explaining the historical context and the main reasons why we think the status of evaluation in France has not yet been stabilised, we will present three French “paradoxes”.

One of them is indeed that - in a meta-evaluation perspective - we are able here to reflect on and review *CSE*'s experience, using its case studies and demonstrating how criteria were built and used to assess the quality of a small number of evaluation studies, the piloting of which was under its supervision - at central state level.

The final section tries to establish how this valuable legacy might influence the future developments of standards in France, in the context of renewed institutions and the development of new actors, among which is the French Society of Evaluation (*Société française de l'évaluation-SFE*)³.

1. HISTORICAL AND INSTITUTIONAL BACKGROUND

After the rather short-lived experiment of the late sixties and early seventies, evaluation emerged again on the French political agenda from the late eighties. At that time, inspired by American practice (Monnier, 1992; Spelenhauer, 1998) under the name *Rationalisation des choix budgétaires (RCB)*, evaluation practice was introduced in the Finance ministry and piloted by its strategic economic studies department (*Direction de la prévision*). The rather “scientific” approach that expected to set up a system able to really rationalize all public expenditure decisions was eventually abandoned in 1983 (Perret, in *CSE*, 1993, p. 71; Monnier, 1992). From that time, it has been a constant French feature that evaluation has never been directly related to the budgetary process.

² By *institutionalisation* we mean that significant institutional bases had been set up and that decisive steps had been achieved in the process of social construction of evaluation, as an activity *as such*.

³ The present analysis draws from several sources. One is the analysis of the reports published by the *CSE* (see references). Another source consists of interviews conducted with former members of the *CSE* as well as members of regional bodies in three regions. A series of systematic interviews was also scheduled by a working group of the *Société française de l'évaluation (SFE)* in 2001, in the context of contributing to its board's strategic guidelines for the future. At this time, the author was both member of the group and *SFE*'s secretary general, activities that provided extensive opportunities for ‘inside’ observation.

The seventies' experience - although generally considered as a failure - was certainly not without impact, inasmuch as it contributed to alter public management references, at least among limited state elites' circles. Duran *et alii* (1995, p. 47) rightly record that the 1986 *Commissariat général du Plan*'s so-called "Deleau" report drew its indubitable inspiration from analyses derived from the "limits of the Welfare State", and, as such, not alien to the *RCB* experiment. Deleau *et alii* (1986) insisted on an orthodox and rather strict "cost efficiency" approach. The tone was totally different when Prime Minister Rocard embarked on an initiative to "modernise public management" in France. Rocard's directive encompassed four main orientations. One of them was the transformation of human resources management in the public sector and evaluation was a second (Perret, in *CSE*, 1993, p. 72). Fontaine and Monnier (1999, p. 18) stress the fact that this important symbolical act used a rare window of opportunity of promoting evaluation in a country generally very alien to the concept.

In January 1990, a *Conseil scientifique de l'évaluation* (*CSE*) was set up by presidential decree to advise a cross-departmental body (*Comité interministériel de l'évaluation-CIME*) created at the same time. As a central government unit, *CIME* was in charge of deciding what evaluations were eligible for funding by a specific central fund. From the start, this meant that *only some* evaluations, agreed upon on a cross-departmental level, were going to be at the centre of the stage of "institutionalised" evaluation. At the same time, all shades of studies and research, as well as audits and inspections were being devised and implemented under the freshly popular name of "evaluation". A considerable activity of conferences and reflections was sparked off at this time.

CSE's legal competence encompassed methods and ethics and was supposed to control the quality of the particular *CIME* evaluations (see Article 8 of the 1990 decree).

The French political and administrative system explains why *CSE* was composed of two main sorts of experts: on the one hand, the *Grands Corps* - i.e. top civil service members (either belonging to the National Statistical Institute (*Insee*), audit courts or inspection units) -; and on the other, the academics. Their participation in the *Conseil* was about equal. In its first composition, the *Conseil* also had one private sector member. It was chaired by an internationally known professor of political science, Jean Leca. A few years later, when new members were nominated, top civil servants formed a slight majority.

After a promising start in the early nineties, as the legal selection procedure of evaluations functioned smoothly, the process was gradually stalled. Most evaluation projects were presented and devised before 1995 and from then on, *CIME* was no longer convened. Accordingly *CSE* was sidelined, and the Prime Minister's office abstained from choosing a new president. As departing members completed their mandate, no new members were nominated.

This explains why the body of meta-evaluation, we are able to analyse, only consists of a handful of operations. In the period 1990-96, fewer than twenty evaluation projects were analysed by *CSE*, out of which less than fifteen underwent the complete process of "recommendations" (we analyse thirteen of them, see below). *CSE*'s grasp of evaluation practice in France was thus at the same time limited and centralised. The scarce quantity of evaluations, it was able to assess, is nevertheless in reverse proportion to its prominent importance in terms of establishing norms and standards and contributing to the definition ("institutionalisation") of evaluation practice in France.

In parallel, from the early nineties a number of *conseils régionaux* (regional authorities) embarked on regional programmes of evaluation. Their involvement in evaluation was given a clear impetus through the passing of a new regulation in 1993, which made evaluation

compulsory for the *contrats de plan État-régions*⁴ (Isaia, 1999). Only some regional authorities then embarked on introducing systematic evaluation and set up special bodies or committees - occasionally involving some kind of partnership with regional state administration's representatives (*Pays-de-Loire*, Brittany, for instance). Involved in commissioning evaluations, designing their procedures and steering their processes, these bodies have also developed limited practice in the area of quality assessing. In a handful of cases, their practice drew from *CSE*'s parallel activity of constructing standards and norms. Nevertheless, the contribution of these regional committees has remained secondary and very informal (Spelenhauer, Warin, 2000). Influences from *CSE* on the gradual construction of varying "doctrines" of evaluation were sometimes the result of particular individuals (as in the case of *Pays-de-Loire* and Brittany)⁵.

1.1. Administrative tradition, multi-level governance and an emerging market

The above developments ought to be understood in the particular French institutional context. One of its essential characteristics is a very uncertain approach to "accountability"⁶. As Perret (in *CSE*, 1993, p. 72) rightly observes, the notion in French is rather badly established. A structural feature of the French institutional context is the centrality of the State and the embedded notion that it is endowed with *intérêt général* ("public welfare" is a partially adequate translation). This explains why top civil servants (the *Grands Corps*)⁷ along with academics - who, incidentally, are also top civil servants - were bound to play a central role in the new "institutionalisation" phase from the late eighties. It should also be stressed that central government in France still commands a "quasi-monopoly" in matters of policy analysis and expertise, although of course a significant part of studies is outsourced. In empirical terms, the "quasi-monopoly" was recently described adequately by a special report commissioned by the French Senate, comparing the United States and French situations (Bourdin, 2001), after a senatorial mission to the US⁸. The ministry of Finance and Insee (the National Statistical Institute, which, by the way, is part of it) here play a prominent role, associated with the central audit and control agencies (*Cour des Comptes*, *Conseil d'État*) (Perret, in *CSE*, 1993, p. 76). Political scientists have comprehensively analysed this situation, most particular to France, which has successfully withstood marginal efforts to introduce more pluralism from the seventies on (Jobert, Théret, 1994). Jean Leca, *CSE*'s former president, stressed that, in the early nineties, no specific social actors had emerged to

⁴ *Circulaire* (9 décembre 1993) "relative à la mise en œuvre de la démarche d'évaluation dans les procédures contractuelles (contrats de plan-contrats de ville).

⁵ At the late stage of *CSE*'s life, an important conference was organised in Brittany, where regional evaluation authorities were invited.

⁶ The French legal tradition differs from the Anglo-Saxon and there is no exact translation of "accountability" in French.

⁷ Duran *et alii* (1995, p. 54) note that "The French State has always relied simultaneously upon the figure of the engineer and the legal scholar in order to 'produce' a society in the name of public authority. Historically, this has legitimised the State's superiority over civil society, which at the same time is assumed to be outside State's control [...]. In this manner, actual or potential policy evaluation structures find themselves closely linked to the civil servants who have traditionally been in charge of administrative control".

⁸ Contrary to the United States or United Kingdom and German situations, independent organisations or foundations are practically inexistent. Presently less than four such small organisations exist today (Bourdin, 2001, pp. 18-25).

engage in independent participation in evaluation (Leca, in *CSE*, 1992, p. 2)⁹. This absence of any organised profession leads to an embedded *de facto* eclecticism, in terms of references and standards, which blurs the frontiers between evaluation and other associated activities (research, consulting, audit and control) (*CSE*, 1992, p. 13)¹⁰.

In a situation combining a “quasi-monopoly” of State expertise and the absence of a profession and standards, the driving force was bound to be provided by the demand side of the evaluation market. This demand is pushed by two factors pertaining to the increasing influence of multi-level governance. On the one hand, European Union (EU) level practice and its general “standards” have played an increasing role, notably because EU level programmes all include the explicit implementation of evaluation regulations (European community’s structural funds). In many areas of traditional social policy, the EU is the dominant buyer of evaluation studies. However, on the other hand, in the complex relationship existing between Member States and the EU Commission, complying with formal regulations may lead to very disparate types of studies: we contend that there is, as yet, very little spill over effect from the EU quality assessment practice to the French debate¹¹. Even in matters of evaluation of the structural funds programmes, there is little cumulative knowledge (Barbier, 1999). The consultants’ profession, in this context, appears intrinsically divided. The big international firms are presently hesitating to invest systematically in a new market where profitability is lower than in the private sector, because of the lower returns offered by administrative commissioners. A significant group of mainstream sectoral-specialised consultants - mostly medium size firms - act as evaluators, but only on the basis of their core sectoral business: they consider evaluation more as a particular “circumstance” for some of their studies, than a really specific activity. Evaluation “specialists”, who make evaluation as their core business, are virtually non existent. On top of this, a very large number of small consultants engage in evaluation on an *ad hoc* and often locally determined basis. “Mainstream evaluation” (if such a notion has a sense in the French context), is thus implemented by “evaluators” who either have a limited grasp of the international state of the art, or explicitly consider that there is no reason for acquiring such knowledge and professional experience¹².

⁹ « Le rôle éminent de l’État, et tout particulièrement du gouvernement » [dans l’évaluation, car] « tant que d’autres acteurs sociaux n’ont pas manifesté suffisamment d’intérêt pour que l’évaluation devienne partie intégrante de la vie publique, le gouvernement a une responsabilité particulière. »

¹⁰ « En l’absence de milieu professionnel les ayant intégrées, la diversité des traditions disciplinaires mobilisées par l’évaluation produit un éclectisme de fait, qui conduit à admettre que pour un même type de question, on puisse utiliser une pluralité d’approches sans tenter de les confronter, de les articuler, ou de les référer précisément à des finalités différentes de l’évaluation. »

¹¹ Specific results of assessments within the Commission are not returned to the various evaluation French *milieus*. The European Commission is never in a position – as far as co-financed programmes are concerned – to dismiss a Member State’s choice.

¹² Very typically, as we interviewed him, the chief executive of one of the significant consultant firms among the medium size operators was completely unable to formulate an answer to what evaluation was and could not identify its contribution to the firm’s turnover or cash-flow.

2. WHY EVALUATION HAS NEVER BEEN REALLY INSTITUTIONALISED IN FRANCE

2.1. Conflicts of conceptions and advocacy coalitions in the “Jacobin” context

In the domain of evaluation, the French have a “speciality”: as Duran *et alii* (1995) have noted, they harbour more controversies about the notion of evaluation than actual practice of it. A typical and enduring controversy opposes “managerial” evaluation to “democratic” evaluation. Although there is obvious analytical substance in the distinction, the long lasting opposition verges on the absurd and is to be related to the uneasy institutional context described above. A third paradigm, namely “pluralistic” evaluation has tried to eschew the opposition between “democratic” and “managerial”, with limited success so far.

For its fiercest critics, “managerial” evaluation roughly fits into a more or less “neo-liberal” agenda, trying to pass as politically neutral. Its only purpose is deemed by them to be cost-cutting. The EU Commission, on the one hand, as a non fully-fledged political entity, can be seen as inevitably married to a “managerial” stance, whatever the extensive efforts its members make to anchor it into an accountability based discourse.

On the other hand, “democratic” evaluation is often seen by its proponents as strictly opposed to any “public management” concern and only interesting and valuable inasmuch as its findings are publicly and democratically discussed. In the French context, put at its extreme, this conception of evaluation is linked to a “voluntaristic”¹³ (“Jacobin”, and often lacking substance) stance in politics.

The idea that good management practice and democracy are incompatible, although indeed very strange, is very commonplace in France. One interesting example was recently provided by Raymond Forni, the current president of the French National Assembly, valuing what he thought to be the on-going “voluntaristic” government success, as opposed to the trivial activity of costing programmes¹⁴. However, the only coherent approach to the management/democratic divide is to consider that both dimensions are closely interlinked (Leca, 1997, pp. 2-14; Leca, 1993). Altogether, the management/democracy controversy has been an important factor explaining why it has been so difficult for organisational initiatives in the French evaluation *milieus* and to agree on a set of norms.

Interestingly, the “pluralistic approach”, described by Duran *et alii* (1995) as *à la française* provides some sort of third way out of the sterile debate. Why is it so? The “pluralistic” approach may be interpreted from numerous angles. It can be seen as the quest for a new “methodology” (or design - in the broadest sense of the term), which is based on certain norms of quality but, because of its particular approach to the question of stakeholders, could act as a political innovation in the context of protracted French actors’ hostility to evaluation.

One of its key characteristics is the requirement - mostly for certain wide ranging “policy evaluations” - to establish an *instance* (the *instance* is an institutional body which may take certain, but not automatically all functions among steering, scientific recommendations,

¹³ To try and translate the French *volontariste*, in the sense that if applied “voluntaristically” political decisions are able to shape things and transform society.

¹⁴ He wrote in *Le Monde* (June, 6, 2001) : « Faire de la politique, même dans l’opposition, ne saurait se résumer à mettre en doute des chiffres et des circuits de financement, à se transformer en une petite chambre des comptes ».

accepting reports and terms of reference, etc., depending on the type of actors represented)¹⁵. The *Conseil scientifique de l'évaluation* has, from the start, insisted on the key role of the *instance* to invent "new forms of pluralism" (CSE, 1992, p. 14)¹⁶. The insistence was on pluralist points of view, not on stakeholders' interests' representation. Key actors were decision-makers, field or operational actors and the beneficiaries of the programmes evaluated. If allowed to function on the basis of an "area of autonomy" (Barbier, 1999, p. 376, after Leca¹⁷), separate from the strict subordination to the political and managerial systems, the *instance* may provide an effective forum for collective and co-operative construction of knowledge about the programmes evaluated. It thus may provide both scientific and political legitimacy to evaluation (Lascoumes, Setbon, 1996, pp. 10-11)¹⁸. CSE's first *rapporteur general adjoint*, Trosa (1992, pp 97-101) also stressed that there was a double meaning (political and "technical") to the approach to a distinctive methodology in evaluation, as against traditional audit practice. That evaluation cannot be confused with audit, in the particular French context of the period, was linked by CSE explicitly to what it termed a "crisis of political representation" in France (CSE, 1992)¹⁹.

Ten years on and with CSE's relative "failure", we can only assume that the above controversy between "managerial and democratic" paradigms has not been resolved and that the "pluralistic" approach has remained fragile and weakly established in the context of enduring structural features pertaining to the institutional context - among which the quasi-monopoly of expertise is prominent. In terms of "cognitive analysis" of policies (Muller, 2000; Sabatier, 2000)²⁰, there is little doubt that no powerful coalition has yet emerged in France to promote the pluralistic approach within a new *référentiel* (or paradigm) of public policy making. We will return to that. Nevertheless, what may appear to day as a relative failure of CSE also points to an indubitable success in terms of establishing and using standards in a coherent manner.

¹⁵ Toulemonde (2000, p. 356) has insisted on the original approach to the *instance*, in the French cultural context, as opposed to other national traditions.

¹⁶ « Pour toutes les évaluations qui concernent des politiques complexes ou sensibles, et qui comportent l'utilisation d'informations et de résultats d'études de nature hétérogène, le Conseil scientifique préconise la mise en place d'une instance d'évaluation. Son rôle est de préciser le questionnement, de piloter les études, d'intégrer leurs résultats et de formuler des conclusions et des recommandations éventuelles qui peuvent être tirées de ces résultats. Sa composition doit tenir compte de la pluralité des points de vue légitimes sur les questions posées, pour des raisons qui tiennent plus à un souci d'efficacité que d'une exigence formelle de représentativité [...]. Les formes de ce pluralisme sont à inventer dans chaque situation, compte tenu de la nécessité de faire de l'instance une véritable structure de travail » (p. 14).

¹⁷ Leca (1993, pp. 190-191; 1997, p. 15) has theorized the role of this area of autonomy.

¹⁸ Perret (in CSE, 1993, p. 67) explicitly notes: « Pour que ce travail aboutisse à des conclusions crédibles aux yeux du commanditaire politique de l'évaluation, il doit être mené selon des règles qui lui confèrent une double légitimité, scientifique et politique [...]. Dans le contexte de l'évaluation de programme, l'évaluateur professionnel assume souvent seul le poids de cette légitimation, à charge pour lui de mobiliser une information et de mettre en œuvre des méthodes crédibles aux yeux des utilisateurs de l'évaluation. Dans le cas des évaluations de politiques publiques telles qu'elles sont pratiquées en France, cette double légitimité est assurée par une instance d'évaluation qui précise le questionnement évaluatif, pilote les études, synthétise et interprète leurs résultats ».

¹⁹ « Si l'on admet qu'une partie de ce qu'il est convenu d'appeler la crise de l'État est en fait une crise de légitimité liée à un déficit de justification du bien-fondé des interventions publiques [...], alors l'évaluation apparaît comme une réponse. » (p. 18)

²⁰ This branch of policy analysis insists on the key role of « social representations » for policies to emerge.

2.2. Conseil scientifique de l'évaluation: a "failure" which is also a success for standards

CSE's record, in quantitative terms, might appear as very limited indeed: fewer than twenty completed rounds of recommendations and meta-evaluation of relatively wide ranging cross-departmental policies. *CIME*, the cross-departmental body, only effectively functioned for the first three years of the procedure.

This situation may be related to the particular difficulty of the centralized procedure, as Chemlinsky (explicitly referring to France) noted in her address to the first Conference of the European Evaluation Society in The Hague (Chemlinsky, 1994, p. 3). Some arguments have also stressed that limits exist to the very notion of a high level of rigour of a "scientific regulator", which acted as a "counter power" with few supportive social actors (Lascoumes, Setbon, in *CSE*, 1996, p. 231). *CSE*'s situation was always ambiguous in this respect, taken between a possible interpretation of its function as "scientific" and its objective as a promoter of a neutral *intérêt général*. The ambiguity existed in terms of the very justification of this neutrality, because evaluations involved accountability to the *instance*, and not to the general public (public debate about the evaluations under *CSE*'s procedure has always remained marginal). As social scientists were always perceived as unaccountable to the general public, power games of administration and commissioners were given an important leeway²¹. Despite initial interest from the *Grands corps*, key actors gradually appeared as disinterested from the stakes involved in *CSE*'s practice, which again points to a lack of social actors promoting the new evaluation *référentiel*.

In terms of "advocacy coalitions", *CSE*'s short history was certainly marked by many disadvantageous factors, among which we can note: (i) internal central State struggles and the conflict of conceptions (Spelenhauer, 1998); (ii) the active reluctance of *Grands corps* (*Conseil d'État*, *Cour des Comptes*, *ministère des Finances*, *Insee*); (iii) the marginal interest shown by academics (economics, sociology, political science); (iv) implicit hostility to the possible dangers of formalisation of standards by consultants and management experts.

Trosa (1993, p. 241) has somehow euphemized this assessment, cursorily summing up the main characteristics of *CSE*'s experience in 1993: "abundant solemnity and minimum learning; a complex procedure devoid of incentives for outcomes; a scientific regulation that the administration appears as unable to draw upon".

When in 1999, after several years of forced inactivity, *CSE* was replaced by the *Conseil national de l'évaluation* (*CNE*) under new circumstances and a completely new design²², the key factors we have listed above were all the more present, despite the recent foundation of a *Société française de l'évaluation* (*SFE*) in the same year.

3. A NUMBER OF FRENCH PARADOXES

From this rapid account of the French institutional context of evaluation, we end up with a number of paradoxes.

²¹ This explains why *CSE*'s recommendations (the first recommendation, see below) were more often than not ignored by commissioners (Lascoumes, Setbon, in *CSE*, 1995, p. 31).

²² In the sense of rather downgrading its role, and especially in terms of standards. However, *CNE*'s legal mission still encompasses the duty to assess the quality of evaluations it is in charge with.

The first one, although formulated nearly seven years ago, is still true. Enthusiasm for evaluation still exists in limited circles (“widespread infatuation” coexists with an absence of knowledge of its meaning, Duran *et alii* [1995, p. 45]). This absence of knowledge is continuing and there is little cumulative knowledge building (although individual good practice evaluations serve for societal learning as everywhere)²³.

New definitions and missions were devised for the *Conseil national de l'évaluation (CNE)* in 1998. Debates including leading scientists and European level professionals are still going on, but they seem to always “bump into” the same hurdles, whereas mainstream politicians still display the same preconceptions and fears of evaluation (as was clearly exemplified by those of them invited to a panel of the 2001 *SFE*’ congress, to the understandable puzzlement of Canadian colleagues).

The second paradox is that, because it is demand-led, the market for evaluation keeps expanding in France. It is led by European Union demand and many evaluations conducted under the European commission’s control (or rather co-steering) certainly pertain to the “ritual compliance” type (EU strict demands are continuously by-passed by political arrangements). On the other hand, the movement for devolution in France - notwithstanding the comical, when not tragic, characteristics of devolution to Corsican authorities - certainly account for a continuous and prospective push to the demand side. This demand, on both counts, appears as independent from any significant construction of a profession and a continuing impossible agreement on the very principle of adopting standards and norms of ethics as a valid basis for founding a particular sphere of research and study.

In this context, the *Société française de l'évaluation* (a late creation indeed in European terms) appeared, after first failed attempts, as a significant window of opportunity. However, four years on, its achievements have been limited, in terms of its membership and their respective representativeness of the disparate French evaluation *milieus*, as well as in terms of its ability to steer and implement a collective ethos for these. The market for evaluation is buoyant and it seems to ignore international references altogether - except for marginal consultants. Within the consultant profession, a (relative) struggle for market shares certainly has hardly any chance to develop along the question of the quality of methods or norms. The 1999 “re-launching” of the central State *CNE* thus took place against mainstream indifference at central State level and little interest at local and regional authorities’ level.

The third paradox pertains to *CSE*’s rigorous approach to the methods and its innovative work (the creation of the *instance*, and more generally, the process and design of large scope policy evaluations, close to the pluralistic model, which were able to take on board the “political” dimension). In the particularly adverse circumstances, *CSE* was nevertheless able to actually implement standards (see next section).

These three paradoxes may be set against a fourth one: in other countries where evaluation has long been established, standards for evaluation seem to be used scarcely, although they exist and are the object of abundant literature.

²³ See the ironic and bitter comment by Leca (1997, p. 9) on the number of official reports on evaluation : « Le nombre de rapports officiels produits en France sur l'évaluation ou mentionnant cette vache sacrée (j'en ai compté au moins six en dix ans) est en proportion inverse de l'intérêt que le gouvernement et le Parlement lui manifestent concrètement - la situation est parfois différente au sein de ministères spécifiés et dans les régions ».

4. CSE'S PRACTICE: THE IMPLEMENTATION OF STANDARDS (1991-96)

4.1. The statutory mission

According to its statutory competence, the *Conseil scientifique de l'évaluation* was not only in charge of monitoring the quality and objectivity of those evaluations that were funded by the National Fund for Evaluation Development. It was endowed with the wider mission²⁴ of "promoting the development of evaluation methods and devising an ethics (*déontologie*) in the domain".

As for the centrally funded evaluations, *CSE* was deemed not only to assess particular evaluation projects *ex ante*, but also to produce formal quality assessments and recommendations, once final reports were handed in. This is certainly the point where the French system's original stance lies.

From 1992-93, the *Conseil* subsequently embarked on establishing a set of quality criteria in order to structure its assessment mission and disseminated them among the various evaluation steering committees involved (*instances*). Perret, who was instrumental in this process, explained that there was explicit hope in *CSE* for it to play a similar role to that which professional societies and groupings played abroad (*CSE*, 1993, p. 77).

CSE's founding regulation defined the steps for quality assessment more precisely, by way of a double recommendation (the *deux avis* procedure²⁵). The first recommendation (*premier avis*) was passed on the basis of the initial project proposed by the would-be commissioners. The latter were then addressed specific guidelines at this initial stage to improve the feasibility of the project or any of its features *CSE* deemed necessary. In a limited but significant number of cases, projects were considered not feasible and altogether abandoned. In a more important number of cases, commissioners had to submit a new project before a second favourable recommendation was passed.

From the reviewing of all evaluations analysed by *CSE* at this first stage, a contradictory finding emerges. On the one hand, *CSE*'s initial recommendations appear as strikingly predictive of what the results of the evaluations were to be at the end of the day. Each time serious reservations were introduced in the recommendations (in terms of feasibility, of the steering process, of the construction of the *evaluanda*, etc.), these anticipatory assessments were extensively confirmed *ex post*. This is because, as has already been hinted at in the previous sections, the commissioners involved very seldom abided by the recommendations and went on with their initial projects, which ended up as unfeasible, or effectively displaying the internal defects *CSE* had pointed out in the first place.

4.2. The process of constructing standards: inspiration from abroad combined with the insistence on an extended notion of "methodology"

Seen with hindsight, the body of standards and norms *CSE* was able to establish at the same time drew on existing reference abroad and added a specific touch concerning the object of the final assessment of evaluations, namely, the final synthetic evaluation report. On the

²⁴ Article 8 of the 1990 regulation.

²⁵ [Le CSE] « formule deux avis : le premier porte sur les méthodes et conditions de réalisation des projets d'évaluation [...], le second porte sur la qualité des travaux effectués et est rendu public en même temps que les évaluations elles-mêmes. » (Article 8, *décret* 90-82, January 22, 1990)

other hand, *CSE* stuck to its specific insistence on the *instance* factor, as a crucial part of an extended notion of “methodology” (or design).

Interviews with former *CSE*'s members show that standards were chosen among existing international references. Its members were well aware of JCSEE²⁶ standards. It is difficult to analyse precisely why such and such a set of criteria prevailed. Chance and individual circumstances certainly played a role. To our knowledge, the process that led to their legitimate use by *CSE* and the experts' community to which it subcontracted part of its expertise did not raise difficulties.

Chen's (1990, pp. 57-65) four “fundamental evaluation values” (“responsiveness” - including relevance and timeliness -; “objectivity”; “trustworthiness” and “generalizability”²⁷) were eventually selected and adopted without difficulty and no particular controversy, as far as our interviews show. For the second step of assessment, the actual meta-evaluation one, and except from the “responsiveness” criterion, the stress was thus laid more on “internal qualities” of evaluation (the judgement included in the final report) than on the consequences of its process, or than on its possible utilization. Although utilization was constantly present in the comments included in the recommendations, it never did figure prominently as one of the essential quality standards. An utilization criterion could have been used, emulating for instance the JCSEE standards or other references²⁸. Historically, in the French situation, *CSE* probably assumed that it was more important to upgrade the internal quality of reports in the first place.

What is more important however is the insistence *CSE* kept putting on a fifth criterion. This one is not included in Chen's values and does not figure generally at this rank of the main quality items in international literature, namely the question formulated as the “transparency of methods” (*la transparence des méthodes*). *CSE* (1993, p. 127) described the standard as following: “this standard implies that evaluation should provide its own directions for use and own limits [...]. These attempts to lucidity and reflexivity are all the more indispensable than the first four criteria mentioned point to ideal requirements which cannot always be abided by at the same time or comprehensively”²⁹. Thus, the “utilization facet” of quality was seen prominently as providing the users of evaluation with a rational assessment of the internal coherence of the global judgement provided by the final report. To our interviewees, of these five eventually selected criteria, “generalizability” proved the most difficult to implement, whereas the other ones were functional. In 1996, *CSE* issued a *Petit Guide de l'évaluation (PGE)* within which the five criteria featured (1996, pp. 46-47). Up to now, this guide has remained the only “official” reference³⁰ akin to standards in the French context.

²⁶ Joint Committee on Standards for Educational Evaluation.

²⁷ As translated in French: *utilité-pertinence; fiabilité; objectivité; possibilité de généralisation*.

²⁸ See JCSEE standards: utility; feasibility; propriety; accuracy. By contrast, see also for instance Patton (1997, p. 16) for “technical quality” and “methodological rigour” criteria (validity; reliability; measurability; generalizability).

²⁹ « Outre l'exigence d'un exposé complet et rigoureux des méthodes mises en œuvre, ce standard inclut l'idée que l'évaluation doit expliciter son propre « mode d'emploi » et ses propres limites : positionnement par rapport à d'autres évaluations possibles sur le même sujet, rappel des questions auxquelles il n'est pas ou mal répondu, anticipation des objections possibles, etc. Cet effort de lucidité et de réflexivité est d'autant plus nécessaire que les quatre premiers standards mentionnés résumant des exigences idéales qui ne peuvent toujours être satisfaites complètement ni simultanément [...] Les utilisateurs de l'évaluation doivent être informés des arbitrages”. éventuellement opérés entre ces différentes exigences. » (p. 127)

³⁰ It being published by the official government publishing house, *La documentation Française*. Leca, *CSE*'s president at that period, wrote in his foreword that the *PGE* could be used in a broader domain than within strict *CSE*'s statutory competence. Originally a second volume was scheduled, which was eventually never published.

4.3. The second consultation on quality (second assessment): first stages

From 1991, CSE exposed its doctrine for the second step of assessment (*second avis*). The stress was clear-cut: “Scientific legitimacy” was to be the leading value selected at that stage (CSE, 1992, p. 71)³¹.

A first evaluation was achieved in 1992, about the use of information technologies in the administration (*L’informatique de l’État*). The report was published with both sets of recommendations (*premier* and *second avis*). CSE was happy because it thought the report had an echo in the professional press and also indirect influence on the information technology departments in the administration (CSE, 1993, pp. 17-18). It particularly mentioned that this favourable impact vindicated the adequacy of the evaluation process and particularly the smooth functioning of the *instance*.

From our interviews appears the conclusion that, in the context of its limited influence and resources, CSE’s members were satisfied that this procedure was adequately fulfilling the requirements of its mission in terms of norms and standards.

4.4. Case studies

Detailed assessments were subsequently published in CSE’s yearly reports. A qualitative analysis of these assessments (see Table) is now possible. The range of policies, as well as their scope, is extremely heterogeneous.

The thirteen evaluations analysed

- Information technologies in administration (SCIS)
- Housing benefits (HB)
- Neighbourhood Housing renewal (HR)
- Special services for the groups experiencing social difficulties (AD)
- Deprived teenagers (DT)
- Social life and school rhythms for children (SLRC)
- Economic Rehabilitation Zones (RZ)
- Special humid zones (ZH)
- Social services for public employees (ASEA)
- Struggle against poverty (SAP)
- Five year Act for employment (A5)
- Prevention of natural hazards (PN)
- Energy control (EC)

4.5. The most used criteria

Because CSE’s recommendations are extensively articulated and always specific to the particular evaluation, detailed and miscellaneous, it is impossible to provide a systematic quantitative analysis of the criteria that were most often used. A more holistic view should be taken. The four above criteria (generalizability except) nevertheless constantly figured as

³¹ « Le Conseil a par ailleurs commencé à réfléchir sur la nature et les finalités du second avis : il apparaît au Conseil que cet avis devrait avoir pour objet principal de porter un jugement sur la légitimité scientifique des conclusions de l’évaluation au vu des informations collectées. »

CSE's synthetic judgement, and were *gradually* standardized as a final section to the recommendations³². The general tone of the recommendations was often more critical and restrained than appreciative and positive appraisals - with few exceptions - seldom appeared. How then were the different criteria (including the four synthetic ones) implemented?

“With regard to the procedure” (or methodology in the extended sense used here, the *instance* particularly), practically all evaluations are commented on. The *Conseil* generally registered the differences arising from its initial recommendations (at project stage) and their actual implementation. This section notably served as exemplifying the ideal role of the *instance*, for instance distinguishing it clearly from a forum where conflicting “interests” could be represented (evaluation AD, in *CSE*, 1994, p. 83) and insisting on the clarity of the initial “mandate” given to it (evaluation ZH, in *CSE*, 1995, p. 253) by the appropriate authorities. *CSE* sometimes stated that the existing body was unable to function as a proper *instance*, along the guidelines it had set (evaluation RZ, in *CSE*, 1995, p. 231). In evaluation HR (*CSE*, 1994, p. 50), compliments were addressed to the *instance*, thus providing a key example of best practice, where the *instance* was continuously and coherently involved at all stages of the evaluation³³.

A second group of criteria included the “evaluation criteria” (effectiveness and so on), “specific methods’ relevance”, the *evaluanda*. These are traditional assessments. Rather often - and this is probably due to the institutional French features analysed in the previous sections - evaluations ended up as having failed to define the “policy” evaluated. This is for instance the case for evaluation HB (*CSE*, 1994, p. 38), where contrary to *CSE*'s initial recommendation, the programmes involved were not sufficiently specified; but also evaluation HR (*ibid.*, pp. 60-62), where the criteria to define what were the characteristics of the target group were never specified. The same applied to evaluation SAP (*CSE*, 1996, p. 65) where the notion of “extreme poverty” was not defined clearly for the evaluation of the struggle against poverty, although the distinction between standard and “extreme” poverty was crucial to the evaluation study.

A third significant group deals with the question of “net effects” and “exact causality”, as well as units of comparison. The most significant example here can be taken from the evaluation of the 1993 Five-year Act on employment policy (evaluation A5 in *CSE*, 1997, p. 16 and following). *CSE* ended up noting that, altogether, the evaluation of this Act was impossible in terms of impact on job creation, because of multiple causal relationships³⁴. But less clear-cut instances also abound, like in the case of evaluation DT (*CSE*, 1994, p. 77) where “excluded teenagers” were not compared to mainstream target groups to determine differential effects across different programmes. In evaluation RZ (*CSE*, 1994, p. 233), *CSE* thoroughly contest the conclusions of the report, with regard to economic development effects and job creation³⁵. These discussions appeared all the more important in the French

³² With the exception of the first evaluations assessed.

³³ « L'instance a joué son rôle dans la mise au point de la méthode d'enquête [...], ainsi que dans le lancement et le suivi des groupes locaux d'évaluation [...]. Les conclusions des deux démarches, bien qu'ayant été rédigées en toute indépendance par rapport à l'instance, ont été reprises telles quelles pour être publiées en même temps que le rapport, ce qui signifie que l'instance en approuve la substance. » (p. 50)

³⁴ « L'impossible évaluation de l'effet sur l'emploi » (p. 31). « Les caractéristiques spécifiques de cette période sont encore incomplètement analysées et l'on manque de recul pour distinguer entre phénomènes conjoncturels et éventuelles ruptures de tendance. Cette incertitude accroît la difficulté d'estimer les effets propres de la loi et de leur imputer une part de l'augmentation du 'contenu en emplois de la croissance' récemment constaté. » (*ibid.*)

³⁵ « La logique même du raisonnement [...] conduit à se demander si la cause de l'effet mesuré n'est pas, au delà des dispositifs et de leur mise en œuvre, un ensemble de facteurs parfois totalement exogènes (par exemple, les atouts géographiques dont disposait la zone au départ [...], susceptibles d'avoir amplifié ou contrarié l'effet propre de la

context than the resort to experimental design or even counterfactual methods had not developed extensively.

A fourth group of recommendations touches on the question of “the general causal framework” and on categories (the causative theory). The assessment on evaluation SLRC (*CSE*, 1995, p. 207) appears as rather scathing, where “the absence of any conceptual framework prejudices the evaluation [...]. Moreover, compromising any possible integration of the data collected, this situation leads to asserting incoherent and disparate answers to the multiplicity of questions raised by the evaluation”. In the case of the delivery of services for “excluded” groups, the particular absence of any definition of what was the specific notion of the servicing situation was criticized (evaluation AD [see Table], in *CSE*, 1994, p. 60). How could one evaluate the quality of a service that was not defined strictly, and involved the complex construction of a service relationship involving a plurality of social actors and “street level” agents?

From the reviewing, two more categories of quality assessment emerge as the most frequent.

As can be anticipated from what was said in the previous sections, the main item concerns the integration of diversified data and conclusions, as well as the “transparency of methods”, into a coherent final report. This report, to *CSE*, should crucially be at the same time able to justify its internal linkage between findings and recommendations and to adopt an “auto-reflexive” stance with regard to its limits.

The other main item deals with the “social utility” of the reports and studies: here *CSE* envisages many possible outlets for utilization of the material it was presented with.

Both items, in a way, are key to a definition of what was and what not evaluation, according to *CSE*'s doctrine was.

Concerning the coherence of reports (integration and transparency of methods), *CSE* proved very hard to please. In many cases, the reports' conclusions were presented as unbalanced, their partiality was criticized as well as the insufficient justification of the basis for the final recommendations written by their authors (evaluations HB, ASEA and A5 are typical of that). The theme of “transparency of methods” was present constantly throughout the evaluations assessed, with but two or three exceptions. In the case of evaluation A5 (*CSE*, 1997, p. 43) the *Conseil* admits that the lack of this transparency should not be entirely imputed to the evaluators, overburdened by data and the complexity of the questions³⁶.

Finally, *CSE* provided abundant comment on how the information from the final reports it assessed could be used. Doing this, it ended up very often as judging the reports as “useful products” with all sorts of possible utilization, but which very often did not function as proper evaluation reports. Evaluation DT (*CSE*, 1994, p. 93) for instance was deemed very useful in terms of the importance of data collected, although it did not provide adequate conclusions to evaluate programmes for deprived teenagers. Evaluation AD (*CSE*, 1994, p. 74) was to be used in a “prospective” way to clarify the types of hurdles experienced by the excluded target groups when they try to use public service delivery, but this information was in no case a proper evaluation of the programmes involved. Evaluation ZH (*CSE*, 1995, p. 257) could act as a consistent body for a first diagnosis, and could initiate further evaluation, once objectives were related to the public policy involved. *CSE* very often

politique [...] On ne peut en tirer la conclusion que, si l'on n'avait pas dépensé cet argent, ces emplois n'auraient pas été créés. » (p. 233)

³⁶ « Le manque de clarté qui affecte malgré tout certains développements provient, pour une part, de l'impossibilité de concilier le souci de maintenir le texte dans un volume raisonnable, le nombre et la difficulté des questions traitées et l'abondance des matériaux dont disposait l'instance. » (p. 43)

acknowledged that the evaluations it assessed achieved a first stage of mapping out the problem and assessing the existing information, without being able to confront proper evaluation questions.

4.6. A clear doctrine for evaluation

From the review implemented here, a more or less explicit *CSE*'s doctrine can be reconstructed with hindsight. The very interesting body of evaluations provides a high quality meta-evaluation corpus.

To *CSE*, at the end of the day, the final report of an evaluation was of prominent importance because *only it* could encapsulate synthetic and rationally articulated final judgements. From this assumption derived the importance granted to the coherent structure of this report, and the consequently privileged status of the "transparency of methods" criterion. More broadly this entailed a clear conception of evaluation, which could be tentatively summed up as follows. Evaluation presumes that clear (and contestable) rational relations are established between: (i) data, findings, (ii) their rearrangement within a causal framework comprising explicit theoretical assumptions, (iii) the subsequent production of synthetic conclusions drawn from this process and, (iv) in some cases their final linkage to a contestable set of recommendations to policy makers.

All in all, *CSE* remained very economical with global positive assessments of the evaluations it analysed (maybe two or three evaluations are in this situation). We would then assume that, because it thought this conception of evaluation was not present, or at least only very partially involved in the selected funded operations it processed, *CSE* implicitly or explicitly stated that, whatever their informative content and social utility, proper evaluations were only achieved in a handful of cases. This might provide us with a fifth paradox in the French case: the implementation of standards leading to the evaluations assessed being considered as not able to be satisfactorily benchmarked against international state of the art.

If an understatement is here allowed, such conclusions were certainly not instrumental in providing *CSE* with allies in the French evaluation, audit and inspection *milieus*, in the context of the very strong "politicization" of evaluation (Lascoumes, Setbon, in *CSE*, 1996, p. 233).

5. REGIONAL COMMITTEES' PRACTICE ON STANDARDS

Compared to the highly articulated body of meta-evaluation here reviewed, the contribution from other significant social actors to the establishing of quality and norms has remained marginal in France so far³⁷. A superficial review of four regions (*Pays-de-Loire*, Brittany, *Rhône-Alpes*, and *Nord-Pas-de-Calais*) indicates that their adoption of standards has been a gradual and pragmatic process. It is however difficult to state that these have been effectively and systematically implemented, apart from the adoption of norms for assessing *projects*. No meta-evaluation in the regional authorities is available to our knowledge. Although obviously benefiting from *CSE*'s influence through various channels (personal influence of members of bodies, dissemination of *CSE*'s publications), the regional committees have rather learnt by doing. *Pays-de-Loire* region for instance, initially benefiting from knowledge transfer,

³⁷ The MEANS (*Méthodes d'évaluation des actions de nature structurelle*) programme for the European Commission of course provides a conspicuous and valuable exception.

established norms rather early for the writing of evaluation reports (inspired by *CSE*'s standards). In the course of steering always more numerous evaluations, the regional committee first addressed questions concerning commissioning (the terms of reference) and then turned to addressing *ex ante* "evaluability" and more lately matters concerning the dissemination of evaluation results. On the other hand, *Rhône-Alpes* apparently implicitly constructed a set of references which entailed a gradually more structured view of evaluation's goals and its specificity. Those were dealt with extensively during a special seminar, which took place after five years of activity. In this document, published in 1997, the regional *Rhône-Alpes* body's president insisted on the fact that one important axis of evaluation was to provide substantial elements for the public and democratic debate and, consequently listed the link with elected officials as an implicit quality criterion.

CONCLUSION: WHAT FUTURE FOR STANDARDS IN FRANCE?

Lack of *CSE*'s immediate influence

It is difficult to eschew the conclusion that *CSE*'s immediate and apparent influence has remained limited in French evaluation. Its gradual marginalization and eventual demise militate against this. All sorts of arguments have been upheld to account for this (the complexity of the process, the question of timing and schedules, and the perception of *CSE*'s members as haughty and inaccessible *gardiens du temple* [sacred wardens]). But the main explanation certainly pertains to the institutional factors (see our first sections).

What is remarkable indeed is that, despite this structural situation, *CSE* proved consistent in fulfilling its explicit mission of contributing to the development of methods, acting as producer of a "doctrine", which it built from its own assessment practice (also drawing upon substantial literature from abroad). It was never in a situation of implementing the dissemination of these norms and appreciating systematically how these were or were not used by the various French evaluation *milieus*. One may differ of course with the content of the doctrine (and especially the key insistence on the structuration of the value judgement, in balance with other dimensions such as utilization), but *CSE* can certainly not be taxed for inconsistency. In a context where nearly anything would pass for evaluation (see Leca's definition of the French way further), the task had to be taken on anyway by some actor.

The enlightenment effect: an investment for the future

Certainly *CSE*'s experience can be considered as a very valuable body of knowledge contributing to enlightenment in Carol H. Weiss' sense. Despite the continuing marginality of evaluation practice in France (if assessed against international standards), the core endeavour of establishing standards especially adapted to at least part of the national culture may be considered as having been, if not complete, very significantly achieved. To us, however, this cannot vindicate an optimistic view of a future large scale utilization of this knowledge base. We would hardly concur here with Fontaine and Monnier's (1999, p. 18) optimistic view as to the medium-term potential of evaluation "professionalization" in the French context. Neither do *SFE*'s first four years of existence lead to optimism.

SFE's recent developments and the context from 1999

The *Société française d'évaluation (SFE)* held its founding conference in Marseilles in June 1999. Its objectives included the “development of evaluation methods and techniques as well as the promotion of the respect of ethical and procedural rules in order to foster the quality of evaluations and the appropriate use of their findings”³⁸. In order to implement this broad objective, a working group on “standards and ethics” has been working from late 1999. It first endeavoured to analyse foreign societies’ ethical guidelines and standards, and working contacts were established with correspondents in these societies (in Australasia, Switzerland, Germany, Italy, the United Kingdom and the United States). This further led to presenting our own synthesis of the sister societies’ documents reviewed, and proposing a working framework for dealing with four types of norms, possibly to include within a future *SFE* chart (norms of social utility, of procedural quality, of product quality and ethical guidelines) (Barbier, Perret, 2000). A version of this chart is presently being discussed within *SFE* and it is currently envisaged that *SFE*’s members (evaluators as well as other participants in evaluations, including the *instances*) could declare their acceptance of the charter on a voluntary basis.

The CNE current context

In 1999, after a period of decline of their activity, *CSE* and *CIME* were eventually replaced by a unique body, the *Conseil national de l'évaluation (CNE)*, with different missions and membership. Whereas *CSE* was statutorily in charge of fostering the development of evaluation methods and defining an evaluation ethics (*déontologie*), the new body has the more limited function of defining the programme of funded evaluations and of assessing their quality *ex post* (Article 5 of the 1998 regulation³⁹). *CNE*’s experience is yet too fresh to be assessed as such.

CONCLUDING REMARKS

The situation described in our first sections has not significantly changed in the present period. Typically, Audebrand (1999, p. 109) was able to quote an interview indicating that in 1998, top level *Conseil d'État* members were still arguing that they did not know *what evaluation was*, whilst Audit court members implemented evaluation their own way (Audebrand, 1999, p. 114). *Insee* trade unions and staff convened an important meeting in 2001 under the theme “*Statistique publique, évaluation et démocratie*”, where a very significant number of top civil servants potentially involved in evaluation was present. The proper question of evaluation was not addressed as such, except for a handful of questions from the audience and a declaration presented in a panel by *SFE*’s secretary.

These facts might be considered puzzling from a non-French perspective. They indicate that the key question for French evaluation is neither the lack of a valuable set of standards in France, nor the lack of their significant and consistent testing on the series of evaluations we

³⁸ See Article 3 of its statute book, partly inspired by other European societies’ statutes – and notably by the European Evaluation Society’s one.

³⁹ To our knowledge, *CNE* has not published quality assessments so far.

have reviewed here, a knowledge base (and a *doctrine*) available to all French evaluation's *milieus*.

We would rather think that the situation points to a protracted absence of decent advocacy coalitions for "evaluation". The question remains as to how long the French exception will be sustainable in the international and (above all) European Union context. Certainly the "convergence" thesis, which envisages "different paths to maturity" (Toulemonde, 2000, p. 355) should be questioned in depth. Welfare regimes - of which the public management ethos is a crucial part - are known to be sustainable within their path dependency, whatever new arrangements they are able to take on board.

Sabatier (1998, p. 8) rightly observed that "the goal should be to develop institutions that force them [policy evaluators] to confront their adversaries in forums dominated by professional norms". But so far, and again despite the quality of CSE's cumulated experience, the endeavours to build these institutions have altogether failed. All in all, Leca's (1997) tongue-in-cheek definition of the French way to evaluation still appears as very adequate, although difficult to render adequately into English⁴⁰.

We are finally confronted with our initial paradox: while the French market for evaluation is expanding, and a considerable body of standards exists including their implementation by CSE, the disparate evaluation *milieus* in France have yet to yield an advocacy coalition that would lead to the proper utilization of the passed years' experience.

BIBLIOGRAPHY

AUDEBRAND E., 1999, *L'évaluation en réforme : la relance du dispositif national d'évaluation des politiques publiques*, mémoire de DEA, sous la direction de J.C. Thoenig, ENS Cachan, 128p.

BARBIER J.-C., 1999, "Intergovernmental Evaluation: Balancing Stakeholders' Expectations with Enlightenment Objectives?", *Evaluation*, vol 5 (4) 373-386, Sage, London.

BARBIER J.-C., PERRET B., 2000, *Ethical Guidelines, Process and Product Quality Standards, What For? An SFE (French Evaluation Society) Perspective*, paper presented at the European Evaluation Society Conference, Lausanne, October, 12th -14th.

BOURDIN J., 2001, *Rapport d'information fait au nom de la délégation du Sénat pour la planification sur l'information économique aux États-Unis*, Sénat, session 2000-2001, n° 326, Paris, 382p.

CHEMLINSKY E., 1994, *Where We Stand Today In The Practice of Evaluation: Some Reflections*, paper for the EES First conference, The Hague, Netherlands, mimeo, 22p.

CHEN H.T., 1990, *Theory-driven evaluation*, Sage, London.

⁴⁰ Leca (1997, p. 10) : « Tout se passe comme si le credo officiel de l'évaluation "à la française" comprenait quelques articles auxquels il suffit de se tenir avec la foi du charbonnier : 1/ l'évaluation est nécessaire ; 2/ il y a bien des manières d'en faire ; 3/ tout le monde évalue, des audits aux inspections et aux corps de contrôle, des commissions de bilan aux comités de sages, et bien, mais sans en faire toute une histoire : n'avons nous pas les meilleurs fonctionnaires du monde dont les rapports font autorité ? ; 4/ l'évaluation n'est pas un label nécessaire ni prestigieux (il y a peu d'exceptions dont la direction de l'évaluation et de la prospective de l'Éducation nationale est la plus éclatante) ; 5/ l'évaluation, quelle que soit sa valeur, est inutile quand elle a pris du temps (elle est remise à quelqu'un qui ne l'a pas demandée, celui qui l'avait demandée ayant quitté son poste ; 6/ l'évaluation est utile quand elle répond vite au désir d'un décideur dans l'urgence et lui indique une solution, quand il est aux abois et cherche une issue acceptable ("évaluation de premier secours"), ou quand il a déjà une solution et cherche à la faire passer ("évaluation appoint"), ou, enfin, à la rigueur, quand elle contribue à socialiser des acteurs et des protagonistes incertains, usés ou mécontents et à prévenir une crise ("évaluation apprentissage" ou "aide au moral des troupes") ; 7/ une évaluation publique est peu utile (si elle se rend insignifiante en voulant satisfaire tout le monde) ou peu pertinente (si elle ne satisfait pas) ; 8/ (le dernier mais pas le moindre) seule l'autorité responsable de l'action publique devrait avoir le droit de valider et publier une évaluation (principe de "l'auto-évaluation publique").

- CSE (Conseil scientifique de l'évaluation), 1992, *L'évaluation, de l'expertise à la responsabilité*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1993, *L'évaluation en développement*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1994, *L'évaluation en développement 1993*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1995, *L'évaluation en développement 1994*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1996, *L'évaluation en développement 1995*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1996, *Petit Guide de l'évaluation des politiques publiques*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1997, *L'évaluation en développement 1996*, La documentation Française, Paris.
- CSE (Conseil scientifique de l'évaluation), 1998, *L'évaluation en développement 1997*, La documentation Française, Paris.
- DELEAU M., NIOCHE J.P., PENZ P., POINSARD R., 1986, *Évaluer les politiques publiques, méthodes, déontologie, organisation*, Commissariat général du Plan, La documentation Française, Paris.
- DURAN P., MONNIER E., SMITH A., 1995, "Evaluation à la française, Towards a New Relationship between Social Science and Public Action", *Evaluation*, vol 1 (1): 45-63.
- FONTAINE C., MONNIER E., 1999, "Evaluation in France", *The evaluation atlas*, Inteval, mimeo.
- ISAÏA H., 1999, *L'évaluation des nouveaux contrats de plan État-région (2000-2006)*, LGDJ, Paris.
- JOBERT B., THÉRET B., 1994, « France, la consécration républicaine du néo-libéralisme », in Jobert B., *Le tournant néo-libéral en Europe*, L'Harmattan, Paris, pp. 21-86.
- LASCOUMES P., 1998, « Évaluer l'évaluation », in Kessler M.C., Lascoumes P., Setbon M., Thoenig J.C. (dir.), *Évaluation des politiques publiques*, L'harmattan, Paris, pp. 23-33.
- LASCOUMES P., SETBON M., 1996, « L'évaluation pluraliste des politiques publiques : enjeux, pratiques, produits », note pour le Commissariat général du Plan, miméo, Paris, 17p.
- LECA J., 1993, « Sur le rôle de la connaissance dans la modernisation de l'État et le statut de l'évaluation », *Revue Française d'Administration Publique*, 66, avril-juin, pp. 185-196.
- LECA J., 1997, « L'évaluation comme intervention, sciences sociales et pratiques administratives et politiques », in Actes du colloque "L'évaluation des politiques publiques", GAPP-CNRS-ENS Cachan des 5 et 6 février, miméo, 19p.
- MONNIER E., 1992, *Évaluations de l'action des pouvoirs publics*, Economica, 2è édition, Paris.
- MULLER P., 2000, « L'analyse cognitive des politiques publiques : vers une sociologie politique de l'action publique », *Revue française de Science politique*, vol. 50, n° 2, avril, pp. 189-208.
- PATTON M. Q., 1997, *Utilization-Focused Evaluation*, *The New Century Text*, Sage, Thousand Oaks.
- SABATIER P.A., 1998, "The Political Context of Evaluation Research, an Advocacy Coalition Perspective", in Kessler M.C., Lascoumes P., Setbon M., Thoenig J.C. (dir.), *Évaluation des politiques publiques*, L'Harmattan, Paris, pp. 129-146.
- SABATIER P.A., SCHLAGER E., 2000, « Les approches cognitives des politiques publiques : perspectives américaines », *Revue française de Science politique*, vol. 50, avril, pp. 209-234.
- SPELENHAUER V., 1998, *L'évaluation des politiques publiques, avatar de la planification*, Thèse de doctorat de science politique, Grenoble, 586 p.
- SPELENHAUER V., WARIN P., 2000, « L'évaluation au service des conseils régionaux », *Sociologie du travail*, 42, 245-252.
- TOULEMONDE J., 2000, "Evaluation Culture(s) in Europe : Differences and Convergence between National Policies", *Vierteljahrshefte Zur Wissenschaftsforschung*, 3, pp. 350-357.

TROSA S., 1992, « Le rôle de la méthode dans l'évaluation à travers l'expérience du Conseil scientifique de l'évaluation en France », *Politiques et management public*, 10, 3, septembre, pp. 83-100.

TROSA S., 1993, « Un premier bilan de l'expérience française », *Revue française d'Administration publique*, 66, avril-juin, pp. 241-248.

NUMÉROS DÉJÀ PARUS :

téléchargeables à partir du site
<http://www.cee-recherche.fr>

- N° 22 *La théorie aokienne des institutions à l'épreuve de la loi de 1841 sur le travail des enfants*
HERVE DEFALVARD
mars 2003
- N° 21 *Pourquoi les entreprises évaluent-elles individuellement leurs salariés ?*
PATRICIA CRIFO-TILLET, MARC-ARTHUR DIAYE, NATHALIE GREENAN
février 2003
- N° 20 *Représentation, convention et institution. Des repères pour l'Économie des conventions*
CHRISTIAN BESSY
décembre 2002
- N° 19 *A Survey of the Use of the Term précarité in French Economics and Sociology*
JEAN-CLAUDE BARBIER
novembre 2002
- N° 18 *Is there a Trap with Low Employment and Low Training for Older Workers in France?*
LUC BEHAGHEL
mai 2002
- N° 17 *From Negotiation to Implementation. A Study of the Reduction of Working Time in France (1998-2000)*
JEROME PELISSE
mai 2002
- N° 16 *Paradoxe, dysfonctionnement et illégitimité de la convention financière*
TRISTAN BOYER
avril 2002
- N° 15 *Déstabilisation des marchés internes et gestion des âges sur le marché du travail : quelques pistes*
JEROME GAUTIE
mars 2002
- N° 14 *Métissage, innovation et travail. Un essai à partir de l'étude d'activités artistiques et culturelles*
MARIE-CHRISTINE BUREAU, EMMA MBIA
mars 2002
- N° 13 *Politique de l'emploi et mise au travail sur des « activités utiles à la société »*
BERNARD SIMONIN
février 2002
- N° 12 *Activité réduite : le dispositif d'incitation de l'Unedic est-il incitatif ?*
MARC GURGAND
décembre 2001

- N° 11 *Welfare to Work Policies in Europe. The Current Challenges of Activation Policies*
JEAN-CLAUDE BARBIER
novembre 2001
- N° 10 *Is the Reproduction of Expertise Limited by Tacit Knowledge? The Evolutionary Approach to the Firm Revisited by the Distributed Cognition Perspective*
CHRISTIAN BESSY
octobre 2001
- N° 9 *RMI et revenus du travail : une évaluation des gains financiers à l'emploi*
MARC GURGAND, DAVID MARGOLIS
juin 2001
- N° 8 *Le statut de l'entrepreneuriat artistique et culturel : une question d'économie politique*
MARIE-CHRISTINE BUREAU
avril 2001
- N° 7 *Le travail des femmes en France : trente ans d'évolution des problématiques en sociologie (1970-2000)*
MARTINE LUROL
mars 2001
- N° 6 *Garder et accueillir les enfants : une affaire d'État ?*
MARIE-THERESE LETABLIER, GERALDINE RIEUCAU
janvier 2001
- N° 5 *Le marché du travail des informaticiens médiatisé par les annonces d'offres d'emploi : comparaison France/Grande-Bretagne*
CHRISTIAN BESSY, GUILLEMETTE DE LARQUIER avec la collaboration de MARIE-MADELEINE VENNAT
novembre 2000
- N° 4 *Le travail : norme et signification*
YOLANDE BENARROSH
octobre 2000
- N° 3 *À propos des difficultés de traduction des catégories d'analyse des marchés du travail et des politiques de l'emploi en contexte comparatif européen*
JEAN-CLAUDE BARBIER
septembre 2000
- N° 2 *L'économie des conventions à l'école des institutions*
HERVE DEFALVARD
juillet 2000
- N° 1 *La certification des compétences professionnelles : l'expérience britannique*
CHRISTIAN BESSY
mai 2000

Ce texte a été soumis **au comité éditorial** du CEE, composé de : Gabrielle Balazs, Jean-Claude Barbier, Christian Bessy, Pierre Boisard, Marie-Christine Bureau, Bruno Courault, Jérôme Gautié, Jean-François Germe, Nathalie Greenan, Martine Lurol, Emmanuelle Marchal, Emmanuèle Reynaud, Bernard Simonin, Marie-Madeleine Vennat, Serge Volkoff.